



Type-based complexity analysis for fork processes

Emmanuel Hainry, Jean-Yves Marion, Romain Péchoux

► To cite this version:

Emmanuel Hainry, Jean-Yves Marion, Romain Péchoux. Type-based complexity analysis for fork processes. 16th International Conference on Foundations of Software Science and Computation Structures (FoSSaCS), Mar 2013, Rome, Italy. pp.305-320, 10.1007/978-3-642-37075-5_20 . hal-00755450v2

HAL Id: hal-00755450

<https://inria.hal.science/hal-00755450v2>

Submitted on 4 Oct 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Type-based complexity analysis for fork processes

Emmanuel Hainry, Jean-Yves Marion, and Romain Péchoux

Université de Lorraine and LORIA

{emmanuel.hainry,jean-yves.marion,romain.pechoux}@loria.fr

Abstract. We introduce a type system for concurrent programs described as a parallel imperative language using while-loops and fork/wait instructions, in which processes do not share a global memory, in order to analyze computational complexity. The type system provides an analysis of the data-flow based both on a data ramification principle related to tiering discipline and on secure typed languages. The main result states that well-typed processes characterize exactly the set of functions computable in polynomial space under termination, confluence and lock-freedom assumptions. More precisely, each process computes in polynomial time so that the evaluation of a process may be performed in polynomial time on a parallel model of computation. Type inference of the presented analysis is decidable in linear time provided that basic operator semantics is known.

Keywords: Implicit Computational Complexity, Tiering, Secure Information Flow, Concurrent Programming, PSpace

1 Introduction

We propose a type system for an imperative language with while loops and forks calls, which provides an upper bound on the complexity of a process by controlling its data flow. Threads are created dynamically and they do not share a global memory. Each `fork` call creates a new child process with a distinct identifier (`id`) and duplicates the execution context including the program counter. The parent process keeps the `ids` of its children (but not the converse). Communications between children and their parent are performed through the use of a `wait` instruction that allows a returning child to pass a value to its parent process. The domain of computation is, in essence, the set of strings, on which various admissible operators can be defined. Thus, we are able to use arrays of strings with respect to the typing limitation as in Example 1.

We demonstrate that each subprocess generated by a well-typed process (under some restrictions) runs in polynomial time (Proposition 2) and that the number of process offsprings is bounded by a polynomial (Proposition 3). As a result, the amount of interactions between two processes is also bounded by a polynomial. Thanks to Savitch's Theorem [21], we show that a program runs in polynomial space on a sequential machine (Theorem 2), and conversely

(Theorem 3). This result is expressed in the central Theorem 1. As far as we know, it is the first characterization of FPSpace based on a typed system for an imperative language. The type inference procedure is computable in linear time in the program size (Proposition 1). As a result, applications about the complexity measure of process calculi may be considered. We refer to the conclusion for a discussion about practical issues.

That said, one of our main motivation is to understand the relationship between data flow control and computational complexity. In [17], a type system was introduced for a sequential imperative programming language, characterizing the set of polynomial time computable functions. The idea behind such a typing system is to control the information flow in an execution by enforcing a data tiering discipline. The idea is that data are ramified into tiers in such a way that data at a given tier may only produce information at the same tier or at a lower tier. Thus, the type system prevents the alteration of upper tier data by lower tier data, following the principle of an integrity policy [3]. From this observation, there is only a small step to do in order to devise a type system based on works on type-based information flow analysis. We refer to the survey [20] for further explanations. The type system assigns tiers (which are similar to security levels) to variables. As in [23, 22], the type system prevents information to flow from lower tier variables to higher tier variables by direct or indirect assignments. From a complexity point of view, the main novelty here is that we have to control the information that flows between processes so that the termination of a father process does not depend on the return values of its children. For this, we suggest a three-tier lattice, $\{-1, 0, 1\}$, where tier **1** variables control the while loops, tier **0** variables are working data, and tier **-1** variables are output values. Consequently, we prevent an unsafe declassification due to the termination of a process through the wait/return mechanism. Lastly, it is worth noticing that we establish a non-interference property (Proposition 2), which expresses the relationship between information flow and complexity.

There are several works, which are directly related to our results. On function algebra, a characterization of PSPACE has been provided by ramified recursion with parameter substitutions [14] and by ramified recurrence over functions [15]. For lambda calculus, [7] introduced a typed lambda calculus based on Soft Linear Logic [11], which identifies exactly PSPACE. On functional languages, several characterizations of PSPACE have been provided in [8] consisting in restrictions on data manipulation and on recursion schemes. The completeness proofs of the aforementioned works simulate alternating polynomial time Turing machines [5]. Another approach is the one of Pola [6] which relates complexity and category theory. On imperative languages, an approach consists in analyzing the growth of the data flow by means of a matrix calculus propagating constraints between variables [19, 9, 18]. Thus, Niggel and Wunderlich gave a characterization of PSPACE [19]. On the complexity of message passing languages based on implicit computational complexity, Amadio and Dabrowski show how to obtain an upper bound on instants for synchronous languages by using quasi-interpretation based tools [1, 2]. More recently, Dal Lago *et al* proposed to polynomially bound process

interactions by type systems based on Light Linear Logic [13, 12]. Notice that the fragment of the process calculi is too weak to establish a completeness result. Lastly, the recent work of Madet [16] proposes a multithreaded program with side effects also based on light linear logic which are polynomial.

2 Imperative language with forks

2.1 Syntax of processes

Expressions, instructions, commands and processes are defined by the following grammar, where \mathcal{V} is the set of variables and \mathbb{O} is the set of operators. The size of an expression is $|X| = 1$ and $|op(E_1, \dots, E_n)| = \sum_{i=1}^n |E_i| + 1$. We note $\mathcal{V}(K)$, $K \in \text{Exp} \cup \text{Proc}$ the set of variables occurring in K .

$$\begin{aligned}
 E, E_1, \dots, E_n \in \text{Exp} &::= X \mid op(E_1, \dots, E_n) & X \in \mathcal{V}, op \in \mathbb{O} \\
 I \in \text{Inst} &::= \text{fork}() \mid \text{wait}(E) \\
 C, C' \in \text{Cmd} &::= X := E \mid C; C' \mid \text{while } E \text{ do } C \mid \\
 &\quad \text{skip} \mid X := I \mid \text{if } E \text{ then } C \text{ else } C' \\
 P \in \text{Proc} &::= \text{return } X \mid C; P
 \end{aligned}$$

2.2 Informal semantics

The semantics is similar to the one of C programs and Unix processes. Each process has an id (a pid). When a fork command is executed, a new child process is created, that will run concurrently with its parent process with its own duplicated memory. The parent process knows the child id, whereas a child has no knowledge of its parent process id. A process P is evaluated inside a *configuration*, a triple $(P, \mu)_\rho$ consisting of a process, a *store* μ mapping each variable to a value of the computational domain and a set of ids ρ . In a configuration, the process can be viewed as the code that remains to be executed, and ρ stores the children process id's. At the creation of a child process, the store μ and the program counter are duplicated, and ρ is \emptyset . The main process id is always 1. When a new child is created by a **fork** its id is automatically set to a new integer and it is recorded in the set ρ of the father's configuration. All the configurations corresponding to the main process and its subprocesses are stored inside an *environment* \mathcal{E} , a partial function mapping the process id to a configuration. Processes do not share a global memory. Consequently, the only way for a process to communicate with its children is through the use of a **wait**(E) instruction, which provides a one-way communication. The following program illustrates how to define distinct code that will be executed by the child but not the parent process and conversely. For the parent, X contains the pid of its child, for the child, X contains 0.

```

P:   X := fork ();           // X contains the child pid
      if X > 0 then {         // father's code (X>0)
        Y := wait(X);
        Y := "father"
      }

```

```

} else {                               // child's code (X=0)
    Y := "child"
}
return Y // The father returns "father" and the child returns "child"

```

2.3 Semantics of expressions, configurations and environments

Domain. Let \mathbb{W} be the set of words over a finite alphabet Σ including two symbols **tt** and **ff** that denote truth values true and false. Let ε be the empty word. The length of a word d is denoted $|d|$. As usual, we set $|\varepsilon| = 0$. Define \sqsubseteq as the sub-word relation over \mathbb{W} , by $v \sqsubseteq w$, iff there are u and u' of \mathbb{W} s.t. $w = u.v.u'$, where \cdot is the concatenation. We write \underline{n} to mean the binary word encoding the natural number n .

Store. A *store* μ is a total function from process variables in \mathcal{V} to words in \mathbb{W} . Let $\mu\{X_1 \leftarrow d_1, \dots, X_n \leftarrow d_n\}$, with X_i pairwise distinct, denote the store μ where the value stored by X_i is updated to d_i , for each $i \in \{1, \dots, n\}$. The size of a store μ , denoted $|\mu|$ is defined by $|\mu| = \sum_{X \in \mathcal{V}} |\mu(X)|$.

Configuration. Given a store μ and a process P , the triplet $c = (P, \mu)_\rho$, where ρ is an element of $\mathcal{P}(\mathbb{N})$, is called a *configuration*. Let \perp be a special erased configuration that will be used to replace the content of a configuration once it has terminated: it no longer uses space. The size of a configuration is defined by $|\perp| = 0$ and $|(P, \mu)_\rho| = |\mu| + \sharp\rho$, where $\sharp\rho$ is the cardinality of ρ .

Expressions and configurations. Each operator of arity n is interpreted by a total function $\llbracket op \rrbracket : \mathbb{W}^n \mapsto \mathbb{W}$. The expression evaluation relation \xrightarrow{e} and the sequential command evaluation relation \xrightarrow{c} are described in Figure 1.

$(X, \mu) \xrightarrow{e} \mu(X)$	(Variable)
$(op(E_1, \dots, E_n), \mu) \xrightarrow{e} \llbracket op \rrbracket(d_1, \dots, d_n),$	if $\forall i, (E_i, \mu) \xrightarrow{e} d_i$ (Operator)
$(\text{skip}; P, \mu)_\rho \xrightarrow{c} (P, \mu)_\rho$	(Skip)
$(X := E; P, \mu)_\rho \xrightarrow{c} (P, \mu\{X \leftarrow d\})_\rho$	if $(E, \mu) \xrightarrow{e} d$ (Assign)
$(\text{if } E \text{ then } C_{\text{tt}} \text{ else } C_{\text{ff}}; P, \mu)_\rho \xrightarrow{c} (C_{\text{tt}}; P, \mu)_\rho$	if $(E, \mu) \xrightarrow{e} \text{tt}$ (If _{tt})
$(\text{if } E \text{ then } C_{\text{tt}} \text{ else } C_{\text{ff}}; P, \mu)_\rho \xrightarrow{c} (C_{\text{ff}}; P, \mu)_\rho$	if $(E, \mu) \xrightarrow{e} \text{ff}$ (If _{ff})
$(\text{while } E \text{ do } C; P, \mu)_\rho \xrightarrow{c} (P, \mu)_\rho$	if $(E, \mu) \xrightarrow{e} \text{ff}$ (While _{ff})
$(\text{while } E \text{ do } C; P, \mu)_\rho \xrightarrow{c} (C; \text{while } E \text{ do } C; P, \mu)_\rho$	if $(E, \mu) \xrightarrow{e} \text{tt}$ (While _{tt})

Fig. 1. Small step semantics of expressions and configurations

Environments. An *environment* \mathcal{E} is a partial function from \mathbb{N} to configurations. The domain of \mathcal{E} is denoted $\text{dom}(\mathcal{E})$ and we denote $\sharp\mathcal{E}$ its cardinal when it is finite. We abbreviate $\mathcal{E}(n)$ by \mathcal{E}_n . The size of a finite environment $\|\mathcal{E}\|$ is defined by $\|\mathcal{E}\| = \sum_{i \in \text{dom}(\mathcal{E})} |\mathcal{E}_i|$. The notation $\mathcal{E}[i := c]$ is the environment \mathcal{E}' defined by $\mathcal{E}'(j) = \mathcal{E}(j)$ for all $j \neq i \in \text{dom}(\mathcal{E})$ and $\mathcal{E}'(i) = c$. As usual $\mathcal{E}[i_1 := c_1, \dots, i_k := c_k]$ is a shortcut for $\mathcal{E}[i_1 := c_1] \dots [i_k := c_k]$. The *initial* environment is noted $\mathcal{E}_{\text{init}}[\mathbf{P}, \mu]$ and consists in the main process with no child. That is $\mathcal{E}_{\text{init}}[\mathbf{P}, \mu](1) = (\mathbf{P}, \mu)_\emptyset$ and $\text{dom}(\mathcal{E}_{\text{init}}[\mathbf{P}, \mu]) = \{1\}$. An environment \mathcal{E} is *terminal* if the root process satisfies $\mathcal{E}_1 = (\mathbf{return} \ X, \mu)_\rho$.

Semantics. The transition \rightarrow for process evaluation is provided in Figure 2. The (Fork) rule creates a new configuration, a new process, say of id n , with a new store, and adds it to the environment. The parent process records its child id \underline{n} into the variable X on which the fork instruction has been called and the child id set of the parent is updated to $\rho \cup \{n\}$. Note also that X is assigned to 0 in the child configuration. The (Wait) commands $Z := \mathbf{wait}(E)$ evaluates the expression E to some binary numeral \underline{n} . If the process \underline{n} is a terminating configuration \mathcal{E}_n with $n \in \rho$, then the output value $\mu'(Y)$ is transmitted and stored in the variable Z . Finally, the returning process n is killed by erasing it through the following operation $\mathcal{E}[n := \perp]$. Note that the side condition $n \in \rho$ prevents locks.

$$\begin{aligned} \mathcal{E}[i := c] &\rightarrow \mathcal{E}[i := c'] \quad \text{if } c \xrightarrow{\mathcal{E}} c' & (\text{Conf}) \\ \mathcal{E}[i := (X := \mathbf{fork}(); \mathbf{P}, \mu)_\rho] &\rightarrow \mathcal{E}[i := (\mathbf{P}, \mu\{X \leftarrow \underline{n}\})_{\rho \cup \{n\}}, n := (\mathbf{P}, \mu\{X \leftarrow 0\})_\emptyset] & (\text{Fork}) \\ \text{with } n = \sharp\mathcal{E} + 1 & \\ \mathcal{E}[i := (X := \mathbf{wait}(E); \mathbf{P}, \mu)_\rho] &\rightarrow \mathcal{E}[i := (\mathbf{P}, \mu\{X \leftarrow \mu'(Y)\})_\rho, n := \perp] & (\text{Wait}) \\ \text{if } (E, \mu) \xrightarrow{\mathcal{E}} \underline{n}, n \in \rho \text{ and } \mathcal{E}_n = (\mathbf{return} \ Y, \mu') & \end{aligned}$$

Fig. 2. Semantics of Environments

2.4 Strong normalization, lock-freedom and confluence

Throughout the paper, given a relation \mapsto , let \mapsto^* be the reflexive and transitive closure of \mapsto and let \mapsto^k denote the k -fold self composition of \mapsto . A process \mathbf{P} is *strongly normalizing* if there is no infinite reduction starting from the initial environment $\mathcal{E}_{\text{init}}[\mathbf{P}, \mu]$ through the relation \rightarrow , for any store μ . Given an initial environment $\mathcal{E}_{\text{init}}[\mathbf{P}, \mu]$, for some strongly normalizing process \mathbf{P} and some store μ , if $\mathcal{E}_{\text{init}}[\mathbf{P}, \mu] \rightarrow^* \mathcal{E}'$, for some environment \mathcal{E}' such that there is no environment \mathcal{E}'' , $\mathcal{E}' \rightarrow \mathcal{E}''$, then either \mathcal{E}' is terminal, i.e. $\mathcal{E}'_1 = (\mathbf{return} \ X, \mu')_\rho$ (the *main process is returning*) or $\mathcal{E}'_1 = (X := \mathbf{wait}(E); C', \mu')_\rho$ (we say that the environment \mathcal{E}' is locked). A process $\mathbf{P} = C; \mathbf{return} \ X$ is *lock-free* if for any initial environment $\mathcal{E}_{\text{init}}[\mathbf{P}, \mu]$, there is no locked environment \mathcal{E}' such

that $\mathcal{E}_{init}[P, \mu] \xrightarrow{*} \mathcal{E}'$. A process P is *confluent* if for each initial environment $\mathcal{E}_{init}[P, \mu]$ and any reductions $\mathcal{E}_{init}[P, \mu] \xrightarrow{*} \mathcal{E}'$ and $\mathcal{E}_{init}[P, \mu] \xrightarrow{*} \mathcal{E}''$ there exists an environment \mathcal{E}^3 such that $\mathcal{E}' \xrightarrow{*} \mathcal{E}^3$ and $\mathcal{E}'' \xrightarrow{*} \mathcal{E}^3$. A strongly normalizing, lock free and confluent process P computes a total function $f : \mathbb{W}^n \rightarrow \mathbb{W}$ defined:

$$\forall d_1, \dots, d_n \in \mathbb{W}, f(d_1, \dots, d_n) = w$$

if $\mathcal{E}_{init}[P, \mu[X_i \leftarrow d_i]] \xrightarrow{*} \mathcal{E}$, for some terminal environment \mathcal{E} with $\mathcal{E}_1 = (\text{return } X, \mu')_\rho$ and $\mu'(X) = w$.

3 Type system

3.1 Tiers and typing environments.

Tiers are elements of the lattice $(\{-1, 0, 1\}, \vee, \wedge)$ where \vee and \wedge are the least upper bound operator and the greatest lower bound operator, respectively. The induced order, denoted \preceq , is such that $-1 \preceq 0 \preceq 1$. In what follows, let α, β, \dots denote tiers in $\{-1, 0, 1\}$. Tiers will be used to type both expressions and commands. Operator types τ are defined by $\tau ::= \alpha \mid \alpha \longrightarrow \tau$. As usual, we will use right associativity for \longrightarrow . A *variable typing environment* Γ maps each variable in \mathcal{V} to a tier. An *operator typing environment* Δ maps each operator op to a set of operator types $\Delta(op)$, where the operator types corresponding to an operator of arity n are of the shape $\tau = \alpha_1 \longrightarrow \dots \longrightarrow \alpha_n \longrightarrow \alpha$.

Intuitively, each tier variable has a specific role to play:

- tier **1** variables will be used as guards of while loops. They should not be allowed to take more than a polynomial number of distinct values;
- tier **0** variables may increase and cannot be used as while loop guards;
- tier **-1** variables will store values returned by child processes and cannot increase. Intuitively, they play the role of an output tape.

3.2 Well-typed processes.

Typing rules. Figure 3 provides the typing rules for expressions, commands and processes. Typing rules consist in judgments of the shape $\Gamma, \Delta \vdash K : \alpha$, $K \in \text{Exp} \cup \text{Cmd}$, meaning that K has type α under variable and operator typing environments Γ and Δ , respectively.

There are some important points to explain in this type system. First, the typing discipline precludes values from flowing from tier α to tier β , whenever $\alpha \preceq \beta$ and $\alpha \neq \beta$. Consequently, the guards of while loops are enforced to be of tier **1** in rule (CW). Moreover, in a (CB) rule, we enforce the tier of the guard to be equal to the tier of both branches. Also note that the subtyping rule (CSub) is restricted to commands in order not to break this preclusion. On the opposite, information may flow from tier **1** to tier **0** then to tier **-1**. This point is underlined by the side condition of the (CA) rule. The (F) rule enforces the tier of the variable storing the process id to be of tier **0** since the value stored

$$\begin{array}{c}
\frac{\Gamma(X) = \alpha}{\Gamma, \Delta \vdash X : \alpha} (EV) \quad \frac{\Gamma, \Delta \vdash E_i : \alpha_i \quad \alpha_1 \longrightarrow \dots \longrightarrow \alpha_n \longrightarrow \alpha \in \Delta(op)}{\Gamma, \Delta \vdash op(E_1, \dots, E_n) : \alpha} (EO) \\
\\
\frac{\Gamma, \Delta \vdash X : \mathbf{0}}{\Gamma, \Delta \vdash X := \text{fork}() : \mathbf{0}} (F) \quad \frac{\Gamma, \Delta \vdash E : \mathbf{0} \quad \Gamma, \Delta \vdash X : -\mathbf{1}}{\Gamma, \Delta \vdash X := \text{wait}(E) : -\mathbf{1}} (W) \\
\\
\frac{\Gamma, \Delta \vdash X : \alpha \quad \Gamma, \Delta \vdash E : \alpha' \quad E \in \text{Exp} \quad \alpha \preceq \alpha'}{\Gamma, \Delta \vdash X := E : \alpha} (CA) \\
\\
\frac{\Gamma, \Delta \vdash C : \alpha \quad \Gamma, \Delta \vdash C' : \alpha'}{\Gamma, \Delta \vdash C; C' : \alpha \vee \alpha'} (CC) \quad \frac{\Gamma, \Delta \vdash E : \mathbf{1} \quad \Gamma, \Delta \vdash C : \alpha}{\Gamma, \Delta \vdash \text{while}(E)\text{do}\{C\} : \mathbf{1}} (CW) \\
\\
\frac{\Gamma, \Delta \vdash E : \alpha \quad \Gamma, \Delta \vdash C : \alpha \quad \Gamma, \Delta \vdash C' : \alpha}{\Gamma, \Delta \vdash \text{if } E \text{ then } C \text{ else } C' : \alpha} (CB) \quad \frac{}{\Gamma, \Delta \vdash \text{skip} : \alpha} (CS) \\
\\
\frac{\Gamma, \Delta \vdash C : \alpha}{\Gamma, \Delta \vdash C : \beta} \alpha \preceq \beta \quad (CSub) \quad \frac{\Gamma, \Delta \vdash X : \beta}{\Gamma, \Delta \vdash \text{return } X : \beta} (R)
\end{array}$$

Fig. 3. Type system for expressions, instructions, commands and processes

will increase dynamically during the process execution. Finally, the tier of the variable storing the result returned by a child process (rule (W)) has to be of tier $-\mathbf{1}$, which means that no information may flow from a variable of a child process to tier $\mathbf{0}$ and tier $\mathbf{1}$ variables of its parent process.

Notations. In practice, we write $C : \alpha$ to say that C is of type α , and E^α to says that E is of type α under the considered environments.

Example 1. The following example illustrates how we can program a reduce operation used in parallel prefix sum [10]. The problem consists in computing the greatest element among n integers in an array-like structure $(A[0], \dots, A[n-1])$.

```

max_reduce( $n^1, A^0$ ) ::=
   $r^0 := 0 : \mathbf{0};$ 
   $f^{-1} := A[r]^0 : -\mathbf{1};$ 
   $\text{flag}^0 := \text{tt} : \mathbf{0};$ 
  while ( $n^1 \neq 1$ )1 do {
    if  $\text{flag}^0$  then {
       $\text{pidl}^0 := \text{fork}() : \mathbf{0}$ 
      if ( $\text{pidl} > 0$ )0 then {
         $r^0 := 2*r+2 : \mathbf{0};$ 
         $\text{pidr}^0 := \text{fork}() : \mathbf{0}$ 
        else {  $r^0 := 2*r+1 : \mathbf{0}$  }
      }
      if ( $\text{pidr} == 0$ )0 or ( $\text{pidl} == 0$ )0 then {  $f^{-1} := A[r]^0 : \mathbf{0};$  }
      else {
         $\text{flag}^0 := \text{ff} : \mathbf{0};$ 
         $\text{xl}^{-1} := \text{wait}(\text{pidl}) : \mathbf{0};$ 
         $\text{xr}^{-1} := \text{wait}(\text{pidr}) : \mathbf{0};$ 
      }
    }
  }

```

// not finished
// father process
// left son
// father


```

      f-1 := max(f-1, max(xl, xr)): 0; } }
    n1 := half(n)1: 1 } // end of while
  return f:-1

```

The notation $A[r]$ simply denotes the access to cell r of the array A .

The program queries the first element of the array $A[0]$ then will spawn two children subprocesses, each one exploring one half of the array (the left son will compute the max of $\{A[i]; 2^k - 1 \leq i \leq 2^k + 2^{k-1} - 2, k \geq 1\} = \{A[1], A[3], A[4], A[7], \dots\}$, the right son computing the max of $\{A[i]; 2^k + 2^{k-1} - 1 \leq i \leq 2^{k+1} - 2, k \geq 1\} = \{A[2], A[5], A[6], A[11], \dots\}$). When labelling the nodes of the subprocesses tree with the character each process checks by itself, it is such that the array is a breadth first search of it.

Notice that the typing of commands uses the subtyping rule *CSub*.

We find convenient to describe now the nature of operators used, that will be presented in the next section (Definitions 1 and 2). The get operation on arrays, expressed above by the assignment $f := A[r]$, is a neutral operation of type $0 \rightarrow 0 \rightarrow 0$ and it can be in *Ntr* because of its type. The length n and the index r are identified by their binary notations. So the operators $2 * r + 2$ and $2 * r + 1$ increase the length by 1 and are positive operators of type $0, 0 \rightarrow 0$, and so are in *Pos*. Similarly, the operator $\text{half}(n)$ divides the length n by two, that is it deletes a letter. So, it is a neutral operator of type $1 \rightarrow 1$ and is in *Ntr*. The predicates or, \neq are typical neutral operators of types resp. $0 \rightarrow 0 \rightarrow 0$ and $1 \rightarrow 1 \rightarrow 1$. However the typing implies that $\text{or} \in \text{Ntr} \cup \text{Pos}$, and $\neq \in \text{Ntr}$. Lastly, max is a max operator of type $-1 \rightarrow -1 \rightarrow -1$ and is in *Max*.

4 Safe processes, type inference and complexity

4.1 Neutral, max, and positive operators

The type system guarantees that information flow goes from tier 1 to tier -1 , and prevents any flow in the other way from a lower tier to higher tier. But this is not sufficient to bound process resources. We need to fix the class of operator interpretations based on their typing. For that we define neutral operators (that make the variables decrease and preserve tier), max operators (that do not make variables grow and preserve lower tiers but not tier 1), and positive operators (that can only produce a result of tier 0 as they can create something bigger than their arguments).

Definition 1.

1. An operator op is neutral if:
 - (a) either op computes a binary predicate (i.e. the codomain of op is $\{\text{tt}, \text{ff}\}$).
 - (b) or $\forall d_1, \dots, d_n, \exists i \in \{1, \dots, n\}, \llbracket op \rrbracket(d_1, \dots, d_n) \leq d_i$.
2. An operator op is max if $\forall (d_i)_{1,n}, \llbracket op \rrbracket(d_1, \dots, d_n) \leq \max_{i \in [1,n]} |d_i|$.
3. An operator op is positive if $\forall (d_i)_{1,n}, \llbracket op \rrbracket(d_1, \dots, d_n) \leq \max_{i \in [1,n]} |d_i| + c$, for a constant $c \geq 0$.

Say that a type $\alpha_1 \longrightarrow \dots \longrightarrow \alpha_n \longrightarrow \alpha$ is decreasing if $\alpha \preceq \wedge_{i=1,n} \alpha_i$. We now give a partition of operators into three classes which depend both on their types and on their growth rates.

Definition 2 (Safe operator typing environment). *An operator typing environment Δ is safe if each type given by Δ is decreasing and there exist three disjoint classes of operators Ntr , Max and Pos such that for any operator op and $\forall \alpha_1 \longrightarrow \dots \longrightarrow \alpha_n \longrightarrow \alpha \in \Delta(op)$, the following conditions hold:*

- If $op \in Ntr$ then op is a neutral operator.
- If $op \in Max$ then op is a max operator and $\alpha \neq \mathbf{1}$.
- If $op \in Pos$ then op is a positive operator and $\alpha = \mathbf{0}$.

Intuitively, expressions in while guards are of tier **1** and so the iteration length just depends on the number of possible tier **1** configurations. Inside a while loop, we can perform operations on variables of other tiers. The tier $-\mathbf{1}$ values are return values and processes are confined in the sense that the information flow of a process does not depend on a return value.

4.2 Main result

Proposition 1 (Type inference). *Given a safe operator typing environment Δ , deciding if there exists a variable typing environment Γ such that the typing rules are satisfied can be done in time linear in the size of the program.*

Proof. We encode the tier of each variable X by 3 boolean variables x_1 , x_0 and x_{-1} . We enforce that variable X is of exactly one tier will be encoded by $(\neg x_0 \vee \neg x_1) \wedge (\neg x_0 \vee \neg x_{-1}) \wedge (\neg x_1 \vee \neg x_{-1})$. Each command gives some constraints. Thus, in the case of an assignment $X := op(Y)$, we need to encode $\Gamma(Y) \geq \Gamma(X)$, which can be represented as $(\neg y_{-1} \vee x_{-1}) \wedge (\neg y_0 \vee \neg x_1) \wedge (\neg x_1 \vee y_1)$. As a result, the type inference problem is reduced to 2-SAT.

Definition 3 (Safe process). *Given Γ a variable typing environment and Δ an operator typing environment, we say that a process P is a safe process if:*

- P is well-typed with respect to Γ and Δ , i.e. $\Gamma, \Delta \vdash P : \beta$;
- Δ is safe.

The main result below is a consequence of the soundness Theorem 2 and the completeness Theorem 3

Theorem 1. *The set of polynomial space computable functions is exactly the set of functions computed by strongly normalizing, lock-free, confluent and safe processes, where a unit-cost is taken as the cost of an operator computation.*

Therefore, the process in example 1 can run within a polynomial space. It is worth noticing that the demonstration below says more about process interaction and runtime, which are, in fact, polynomially bounded as we shall see.

5 Complexity soundness

5.1 Process runtime

We establish that each process runs in polynomial time if the time measure is the number of reductions. We follow the line of the soundness proof of [23] to demonstrate a non-interference property. We first prove two lemmas 1 and 2 in order to express that an expression of tier **1** just depends on tier **1** variables, and not on lower tier variables. As a result, there is only a polynomial number of tier **1** configurations, because of neutral operator growth rate. From the security point of view, this means that a variable of tier (level) **1** can be updated by information of the same tier of integrity. Proposition 2 corresponds to the non-interference property. Intuitively, it says that the complexity of a process depends only on tier **1** variables, and so it is not modified if values of lower tiers are updated and so the process runtime does not interfere with lower tier values.

Lemma 1 (Simple security). *Given a safe process P wrt typing environments Γ and Δ , if $\Gamma, \Delta \vdash E : \mathbf{1}$, for an expression E in P , then for each $X \in \mathcal{V}(E)$, $\Gamma(X) = \mathbf{1}$ and all operators in E are neutral.*

Proof. By induction on E .

Given a typing environment Γ say that $|\mu|_i = \sum_{\Gamma(X)=i} |\mu(X)|$. We will first prove that the values taken by each tier **1** expression are at most in polynomial number.

Lemma 2. *Given a safe process P wrt typing environments Γ and Δ , for each expression E in P such that $\Gamma, \Delta \vdash E : \mathbf{1}$, the number of distinct values taken by E during the evaluation of the initial environment $\mathcal{E}_{init}[P, \mu]$ is bounded polynomially in $|\mu|_1$.*

Proof. A store can be updated either (i) by using a **fork** instruction, (ii) by using a **wait** instruction or (iii) by assigning the result of an expression. The typing discipline prevents variables assigned to in (i) and (ii) to be of tier **1**. For (iii), we have $E = op(E_1, \dots, E_n)$, for some neutral operator op , by Lemma 1. Consequently, the result of the computation is either **tt** or **ff** or a subword of the initial values. The number of sub-words being quadratic in the size, the number of distinct values of tier **1** variables is at most quadratic in $|\mu|_1$.

The relations \Rightarrow_i are defined to express the computation of the process of id i .

Definition 4 (\Rightarrow_i). *Given two environments \mathcal{E} and \mathcal{E}' such that $\mathcal{E} \rightarrow \mathcal{E}'$, we write $\mathcal{E} \rightarrow_i \mathcal{E}'$ if the transition \rightarrow corresponds to the evaluation of the process in the configuration \mathcal{E}_i of \mathcal{E} . In the same way define the transition $\mathcal{E} \rightarrow_{\neq i} \mathcal{E}'$ if there exists $k \neq i$ such that $\mathcal{E} \rightarrow_k \mathcal{E}'$. Now define \Rightarrow_i by $\mathcal{E} \Rightarrow_i \mathcal{E}'$ if there are environments $\mathcal{E}^1, \mathcal{E}^2$ such that $\mathcal{E} \rightarrow_{\neq i}^* \mathcal{E}^1 \rightarrow_i \mathcal{E}^2 \rightarrow_{\neq i}^* \mathcal{E}'$.*

The proposition below expresses that the number of instructions performed by a single process is bounded by a polynomial in the tier **1** initial values. Intuitively, this means that each process runs in polynomial time, in the tier **1** initial values, if we do not count the waiting time due to forks.

Proposition 2 (Non-interference wrt time). *Given a strongly normalizing and safe process P , there is a polynomial Q such that, for each initial environment $\mathcal{E}_{init}[P, \mu]$, $\forall i \in \mathbb{N}$, if $\mathcal{E}_{init}[P, \mu] \Rightarrow_i^k \mathcal{E}$ then $k \leq Q(|\mu|_1)$.*

Proof. The typing discipline enforces all the expressions in while-loop guards of a safe process to be of tier **1**. The evaluation of tier **1** values does not depend on lower tier values. By Lemma 2, the number of values taken by tier **1** variables during the evaluation is bounded by $Q(|\mu|_1)$ for some polynomial Q . If a process enters twice in the same configuration with the same tier **1** values, then the computation loops forever. Consequently, if P is strongly normalizing then all the while loops are executed at most $Q(|\mu|_1)$ times.

5.2 Process spawning

Now, we demonstrate that given a strongly normalizing and safe process the number of children of a process (Lemma 3), the number of process generations (Lemma 4) and the size of a configuration (Lemma 5) are polynomially bounded.

Definition 5. *Given a finite environment \mathcal{E} , the process tree $T(\mathcal{E})$ is defined:*

- the nodes are the configurations $\{\mathcal{E}_1, \dots, \mathcal{E}_{\# \mathcal{E}}\}$ and the root is \mathcal{E}_1 ;
- for each $l \in [1, \# \mathcal{E}]$, there is an edge from $\mathcal{E}_l = (P, \mu)_\rho$ to \mathcal{E}_k , if $k \in \rho$.

The degree $d(T)$ corresponds to the number of children generated by fork instructions of a given process. The height $h(T)$ is the number of nested processes.

Lemma 3. *Given a strongly normalizing and safe process P , there exists a polynomial Q such that, for each initial environment $\mathcal{E}_{init}[P, \mu]$, if $\mathcal{E}_{init}[P, \mu] \rightarrow^* \mathcal{E}$ then $d(T(\mathcal{E})) \leq Q(|\mu|_1)$. In other words, for each $\mathcal{E}_i = (P_i, \mu_i)_{\rho_i}$, $i \leq \# \mathcal{E}$, the number of subprocesses is bounded by $Q(|\mu|_1)$, i.e. $\# \rho_i \leq Q(|\mu|_1)$.*

Proof. By Proposition 2, there is a polynomial Q such that the transition \rightarrow_i is taken at most $Q(|\mu|_1)$ times, which bounds the number of executed **fork** instructions.

Lemma 4. *Given a strongly normalizing and safe process P , there exists a polynomial Q such that, for each initial environment $\mathcal{E}_{init}[P, \mu]$, if $\mathcal{E}_{init}[P, \mu] \rightarrow^* \mathcal{E}$ then $h(T(\mathcal{E})) \leq Q(|\mu|_1)$.*

Proof. By Proposition 2, there is a polynomial Q which bounds the number of executed instructions and which just depends on the size of tier **1** initial values. Now, when $X := \text{fork}()$ is performed, the parent process store is duplicated in the child process store, except for the value of X . But X is of tier 0 and so has no impact on the computational time of both processes. Next, a fork command has been executed in the parent process. So, the runtime of the child generated is strictly less than the runtime of its father. Therefore, we deduce that Q bounds the height of the process tree.

We end by showing that subprocess stores are polynomially bounded in the size of the initial store, which is a consequence of the non-interference property, as stated in Proposition 2.

Lemma 5. *Given a strongly normalizing and safe process P , there exists a polynomial S such that, for each initial environment $\mathcal{E}_{init}[P, \mu]$, if $\mathcal{E}_{init}[P, \mu] \rightarrow^* \mathcal{E}$ then $\forall i \leq \sharp \mathcal{E}$, if $\mathcal{E}_i = (P_i, \mu_i)_{\rho_i}$ then $|\mu_i| \leq S(|\mu|_1) + |\mu|_0 + |\mu|_{-1}$.*

Proof. There are three cases to consider. First, a variable of tier **1** is updated by an expression of tier **1**. By Lemma 1, a tier **1** expression just consists in neutral operators and tier **1** variables. So, tier **1** variables are always sub-words of tier **1** initial values. Thus, the size of a variable of tier **1** is always bounded by $|\mu|_1$. Second, take a tier **0** variable X , which is either updated inside a process by a composition of operators or updated by a pid return of a fork. By combining Proposition 2 and lemmas 3 and 4, we obtain a polynomial R which depends on operators and on the program such that the size of X is bounded by $R(Q(|\mu|_1)) + |\mu|_0$ in each process computation. Third, take a tier **-1** variable Y . Suppose that the variable Y is assigned to max operators and variables of any tier. Observe that the case when $Y := \text{wait}(E')$ is a particular case of an assignment if we see globally all processes together. Hence Y is bounded by $R(Q(|\mu|_1) + |\mu|_0 + |\mu|_{-1})$.

5.3 PSpace abiding evaluation strategy

We define a deterministic evaluation strategy \rightarrow starting from the initial environment and evaluating it until it reaches a wait instruction for a process n . Then, the strategy runs the process n until it returns a value. Formally, define a state $s = (\mathcal{E}, l)$ to be a pair of an environment \mathcal{E} and a stack l of ids representing the queued processes. The initial state is $(\mathcal{E}_{init}[P, \mu], [1])$. We denote $::$ the stack constructor.

1. If $\mathcal{E} \rightarrow_h \mathcal{E}'$ for some rule (R) of Figure 2, $R \neq \text{Wait}$, then $(\mathcal{E}, h :: q) \rightarrow (\mathcal{E}', h :: q)$
2. If $\mathcal{E}_h = (X := \text{wait}(E); P, \mu)_\rho$ and $(E, \mu) \xrightarrow{e} \underline{n}$ then $(\mathcal{E}, h :: q) \rightarrow (\mathcal{E}, n :: h :: q)$
3. If $\mathcal{E}_n = (\text{return } X, \mu)_\rho$ and $\mathcal{E}_h = (Y := \text{wait}(E); P, \mu')_{\rho'}$ then $(\mathcal{E}, n :: h :: q) \rightarrow (\mathcal{E}', h :: q)$ where $\mathcal{E} \rightarrow_h \mathcal{E}'$ for the (Wait) rule of Figure 2.

Notice that the rule (2) implies that $(E, \mu') \xrightarrow{e} \underline{n}$ in rule (3).

Lemma 6 (Correction Lemma). *Given a strongly normalizing and confluent process P and an initial environment $\mathcal{E}_{init}[P, \mu]$, if there exists \mathcal{E} such that $\mathcal{E}_{init}[P, \mu] \rightarrow^* \mathcal{E}[1 := (\text{return } X, \mu')_\rho]$ then there exists \mathcal{E}' s.t. $(\mathcal{E}_{init}[P, \mu], [1]) \rightarrow^* (\mathcal{E}'[1 := (\text{return } X, \mu'')_{\rho'}], [1])$ and $\mu''(X) = \mu'(X)$, where $[1]$ is the stack containing the single process 1.*

The size of a stack $|l|$ is the number of elements in it (e.g. $|[1]| = 1$).

Lemma 7. *Given a strongly normalizing and safe process P , there is a polynomial Q such that for each initial environment $\mathcal{E}_{init}[P, \mu]$ and each stack l , if $(\mathcal{E}_{init}[P, \mu], [1]) \rightarrow^* (\mathcal{E}, l)$ then (i) $|l| \leq Q(|\mu|)$ and (ii) $\|\mathcal{E}\| \leq Q(|\mu|)$.*

Proof. The \rightarrow strategy explores the process tree using a stack. The height of the stack corresponds to the number of nested processes. The first inequality follows Lemma 4. Next, by combining Lemmata 3 and 5, we obtain the second inequality.

Theorem 2 (Soundness). *A strongly normalizing, confluent and safe process P can be evaluated in polynomially bounded space using strategy \rightarrow , where a unit-cost is taken as the cost of an operator computation.*

6 Completeness

Let us show that each polynomial space computable function can be computed by a strongly normalizing and safe process.

Theorem 3. *Every polynomial space function is computable by a strongly normalizing, lock-free and safe process P .*

Proof. We present how to compute the value of a quantified boolean formula (QBF). The program below may be seen as a skeleton from which we may simulate, for example, an alternating Turing machine running in polynomial time. Consequently, the computation of a polynomial space function necessitates to compute its output bit by bit, which may be uniformly performed by generating each address and querying the output bit. The program below generates 2^n forks where n is the number of variables. We suppose that the concrete syntax of a QBF is implemented as a string `phi` of tier **1**. Two neutral operators `kind` and `variable` of type **1** \rightarrow **1** return the quantifier kind and the variable bound at the root of `phi`. The neutral operator `next` of type **1** \rightarrow **1** erases the root quantifier and moves to the next one. The positive operator `evaluate` of type **1, 0** \rightarrow **0** computes a boolean formula with respect to an evaluation encoded as an array of booleans, which is an NC¹ complete problem [4]. As in example 1, arrays are words of type **0**, whose length is bounded by `phi`. The set array operator, that we conveniently note `vartab[x] := tt`, is a positive operator of type **0, 0** \rightarrow **0**. We consider the array to be pre-allocated using a `calloc` operator that can be implemented using a positive operator in a loop.

This program has two big while loops: the first one generates 2^n subprocesses, one for each possible variable configuration, for each variable, the father will assign `tt` and the son will assign `ff`; the second one gathers the results of the children, making a disjunction with the parent if they correspond to a \exists , a conjunction for a \forall .

```
qbf(phi) ::=
  psi1 := phi1 : 1;
  q1 := kind(phi1) : 1;
```

```

x1 := variable(phi1) : 1;
i0 := 00 : 0; // son number of a process
pidtab0 := calloc(phi1, phi1) : 0;
vartab0 := calloc(phi1, tt) : 0;
while (q1 == '∃') or (q1 == '∀') do {
  phi1 := next(phi1) : 1;
  pid0 := fork() : 0;
  if pid0 > 0 then {                                     // father process
    pidtab[i] := (pid, q) : 0;
    vartab[x] := tt : 0;
    i0 := i + 10 : 0
  } else {                                              // son process
    vartab[x] := ff : 0;
    i0 := 00 : 0 } // end of else
  q1 := kind(phi1) : 1;
  x1 := variable(phi1) : 1; } // end of while
res-1 := evaluate(phi1, vartab)0 : 0;
while (state(psi1) == '∃') or (state(psi1) == '∀') do {
  psi1 := next(psi1) : 1;
  if (i > 0) then { // for each son
    i0 := i - 10 : 0;
    (pid0, op0) := pidtab[i] : 0;
    res_son-1 := wait(pid0)-1 : -1
    if op0 == '∃' then
      res-1 := or(res, res_son)-1 : 0;
    else
      res-1 := and(res, res_son)-1 : 0; } }
return res-1

```

7 Conclusion

We established that a typed process will generate only a polynomial number of offspring processes and that each process runs within polynomial time bounds. This work may have practical applications to determine whether or not a process runs within certain limited computing resources. Indeed, it may be a tool to control the spawning mechanism, and so to prevent, for example, denial of service attacks.

Let us come back to Example 1 describing a max-reduce algorithm. The runtime is $O(\log^2(n))$. The reason of this apparent shortcoming is that we can not allow a break instruction inside a while-loop. In this case, a return instruction would "disrupt" the information flow and would interfere with tier 1 values. As a result, the runtime would not be anymore bounded. However, if the goal is to analyze program complexity, then we can devise a program transformation Θ , which, approximately, moves break instructions outside while loops, whenever it is possible. Such program transformation may be efficiently performed by a tree transducer. Moreover, there is no reason to preserve the program semantics because we are just interested in resource bounds. So, we

may omit implementation details, which may increase drastically the expressivity of the considered programs. Therefore, we may think of this overall scenario: A system receives a process P to run. First, it computes an abstraction of P using a program transformation Θ . Second, it checks that the abstraction is well-typed. As a result, and even if the system does not know precisely an upper-bound on the complexity, it gets efficiently some confidence on the resource usage of P .

The program solving QBF gives another limitation because we are not allowed to make a loop bounded by the number of subprocesses, which is a tier **0** value. However, we know that this number is bounded by a tier **1** value. To solve this, Hofmann suggested to manage the garbage collector thanks to a type system. We may think to allow iterations on controlled unsafe values by a kind of declassification rule.

The above questions suggest the need to delve deeper into the question of determining the amount of information, which can be declassified while guarantying complexity. In regards to future research, the relationship between information, information flow control and complexity should be better understood.

References

1. R. M. Amadio and F. Dabrowski. Feasible reactivity for synchronous cooperative threads. *Electron. Notes Theor. Comput. Sci.*, 154(3):33–43, 2006.
2. R. M. Amadio and F. Dabrowski. Feasible reactivity in a synchronous pi-calculus. In *Proceedings of the 9th ACM SIGPLAN international conference on Principles and practice of declarative programming*, PPDP '07, pages 221–230, New York, NY, USA, 2007. ACM.
3. K. Biba. Integrity considerations for secure computer systems. Technical report, Mitre corp Rep., 1977.
4. S. R. Buss. The boolean formula value problem is in ALOGTIME. In *Proceedings of the nineteenth annual ACM symposium on Theory of computing*, STOC '87, pages 123–131, New York, NY, USA, 1987. ACM.
5. A. Chandra, D. Kozen, and L. Stockmeyer. Alternation. *J. ACM*, 28(1):114–133, 1981.
6. R. Cockett and B. Redmond. A categorical setting for lower complexity. *Electron. Notes Theor. Comput. Sci.*, 265:277–300, 2010.
7. M. Gaboardi, J.-Y. Marion, and S. Ronchi Della Rocca. A logical account of PSPACE. In *POPL '08*, pages 121–131. ACM, 2008.
8. N. D. Jones. The expressive power of higher-order types or, life without cons. *J. Funct. Program.*, 11(1):5–94, 2001.
9. N. D. Jones and L. Kristiansen. A flow calculus of *wp*-bounds for complexity analysis. *ACM Trans. Comput. Log.*, 10(4), 2009.
10. Richard E. Ladner and Michael J. Fischer. Parallel prefix computation. *J. ACM*, 27(4):831–838, October 1980.
11. Y. Lafont. Soft linear logic and polynomial time. *Theor. Comput. Sci.*, 318(1-2):163–180, 2004.
12. U. Dal Lago and P. Di Giamberardino. Soft session types. In *EXPRESS 2011*, volume 64 of *EPTCS*, pages 59–73, 2011.
13. U. Dal Lago, S. Martini, and D. Sangiorgi. Light logics and higher-order processes. In *EXPRESS'10*, volume 41 of *EPTCS*, pages 46–60, 2010.

14. D. Leivant and J.-Y. Marion. Ramified recurrence and computational complexity II: Substitution and poly-space. In *CSL*, volume 933 of *LNCS*, pages 486–500, 1994.
15. D. Leivant and J.-Y. Marion. Predicative functional recurrence and poly-space. In *TAPSOFT'97*, volume 1214 of *LNCS*, pages 369–380. Springer, 1997.
16. A. Madet. A polynomial time lambda-calculus with multithreading and side effects. In *PPDP*, 2012.
17. J.-Y. Marion. A type system for complexity flow analysis. In *LICS*, pages 123–132, 2011.
18. J.-Y. Moyen. Resource control graphs. *ACM Trans. Comput. Logic*, 10(4):29:1–29:44, 2009.
19. K.-H. Niggl and H. Wunderlich. Certifying polynomial time and linear/polynomial space for imperative programs. *SIAM J. Comput.*, 35(5):1122–1147, 2006.
20. A. Sabelfeld and A. C. Myers. Language-based information-flow security. *IEEE J. Selected Areas in Communications*, 21(1):5–19, 2003.
21. W. J. Savitch. Relationship between nondeterministic and deterministic tape classes. *JCSS*, 4:177–192, 1970.
22. G. Smith and D. Volpano. Secure information flow in a multi-threaded imperative language. In *POPL*, pages 355–364. ACM, 1998.
23. D. Volpano, C. Irvine, and G. Smith. A sound type system for secure flow analysis. *Journal of Computer Security*, 4(2/3):167–188, 1996.